MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD-A160 321

②

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
UNCLASSIFIED	
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
	AFOSR-TR- 85-0774

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
University of Texas at Austin		AFOSR/NM
6c. ADDRESS (City, State and ZIP Code)		7b. ADDRESS (City, State and ZIP Code)
Austin, Texas 78712		Building 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR	NM	AFOSR-81-0047

8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.			
Bldg. 410 Bolling AFB, D.C. 20332-6448	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
	61102F	2304	A5	

11. TITLE (Include Security Classification) SOME COMMENTS ON THE DESIGN OF QUANTIZERS (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
Efren F. Abaya (University of the Philippines) and Gary L. Wise

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Interim	FROM 10/1/80 TO 9/30/85	1985, May 30	8

16. SUPPLEMENTARY NOTATION
PROCEEDINGS OF THE TWENTY-SECOND ANNUAL ALLERTON CONFERENCE ON COMMUNICATION, CONTROL, AND COMPUTING, Monticello, Illinois, October 3-5, 1984, pp. 579-586.

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	quantizers, quantizer design, empirical distribution, design algorithms

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
In many applications involving quantization the probability distribution of the input signal is unknown. However, most of the algorithms for optimal scalar or vector quantization require an explicit distribution function or probability density. This paper shows that under certain conditions reasonable quantizer designs can be expected when standard algorithms are applied to estimates of distribution functions.

DTIC
SELECTE
OCT 15 1985

DTIC FILE COPY

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Brian W. Woodruff, Major, USAF	(202) 767-5027	AFOSR/NM

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.

SOME COMMENTS ON THE DESIGN OF QUANTIZERS

EFREN F. ABAYA
Department of Electrical Engineering
University of the Philippines
Diliman, Quezon City 3004, Philippines

and

GARY L. WISE
Departments of Electrical and Computer Engineering and Mathematics
University of Texas at Austin
Austin, Texas 78712

ABSTRACT

In many applications involving quantization the probability distribution of the input signal is unknown. However, most of the algorithms for optimal scalar or vector quantization require an explicit distribution function or probability density. This paper shows that under certain conditions reasonable quantizer designs can be expected when standard algorithms are applied to estimates of distribution functions.

I. INTRODUCTION

Much work has been expended in studying algorithms for optimal quantization of a known probability distribution [1-5]. In practice, however, the statistical description of the source is rarely known precisely. In some recent papers [2,3] a group of researchers demonstrated both theoretically and experimentally that a training sequence of independent or ergodic samples can be used to design near optimal vector quantizers, by using the sample empirical distribution in place of the unknown input distribution in a generalized version of Lloyd's Method I [1]. They showed that under some conditions the quantizer designed for a "long" training sequence approximates closely the output levels and performance of the optimal quantizer for the true (unknown) distribution. The same kind of reasoning should hold for any design algorithm. If the input distribution F is not known, then we can form an estimate $\hat{F}_n$ based on n observations of the input signal. As n becomes large, we expect a reasonable estimate to converge to the true distribution F. Intuitively, then, an optimal quantizer designed for $\hat{F}_n$, and the resulting distortion, should closely approximate those of an optimal quantizer for F. In this paper we will establish properties of an estimator $\hat{F}_n$ so that this kind of reasoning will be valid.

II. DEVELOPMENT

An N-level k-dimensional vector quantizer is a mapping $Q: \mathbb{R}^k \to \mathbb{R}^k$ which assigns to the input vector x an output vector Q(x) chosen from a finite set of N vectors $\{y_i: y_i \in \mathbb{R}^k, i=1,2,\ldots,N\}$. The distortion incurred in quantizing a k dimensional random variable X having a probability distribution function F is expressed by

$$D(Q,F) = \int C_0(\| x-Q(x)\|)\, dF(x) \qquad (1)$$

where $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^k$ and where all integrals, unless noted otherwise, are over $\mathbb{R}^k$. We will take the cost function $C_0(t)$ to be nonnegative, nondecreasing on $[0,\infty)$ and lower semi-continuous. It has been shown previously that optimal quantizers minimizing (1) exist for

85 10 11 179

all probability distributions F and all k and N [6,7]. This type of distortion function subsumes most of the popular error criteria used in scalar and vector quantization. When optimal quantizers are being considered, there is no loss of generality in assuming the "nearest neighbor" assignment rule: $Q(x)$ is that member of the set which is nearest to x in Euclidean norm, with ties being broken by an arbitrary preassigned method. This rule will be adopted throughout the rest of the paper. Thus a quantizer is completely represented by its set of output vectors.

We will say that a sequence of N-level quantizers $\{Q_n\}$ converges weakly to the quantizer Q if $Q_n(x) \rightarrow Q(x)$ at all continuity points x of Q.

Let F_n and F be k-variate probability distribution functions. The sequence $\{F_n\}$ is said to converge weakly to F (written $F_n \xrightarrow{W} F$) if $F_n(x) \rightarrow F(x)$ at every continuity point x of F. We say that $\{F_n\}$ converges setwise to F (denoted $F_n \xrightarrow{S} F$) if

$$\lim_{n \to \infty} \int_B dF_n(x) = \int_B dF(x)$$

for every Borel subset B of $\mathbb{R}^k$.

The following theorem from [8] and [9] is the main tool used in the investigation. Here k and N are fixed positive integers.

<u>Theorem 1:</u> Assume that $C_0(t)$ is nonnegative and nondecreasing on $[0,\infty)$. Suppose that $C_0(\|x-y\|)$ is uniformly integrable with respect to a sequence of distribution functions $\{F_n\}$ for every y. Let Q_n be an optimal N-level quantizer for F_n. If either (a) $C_0(t)$ is continuous and $F_n \xrightarrow{W} F$; or (b) $C_0(t)$ is lower semicontinuous and $F_n \xrightarrow{S} F$, then every convergent subsequence of $\{Q_n\}$ converges weakly to an optimal quantizer for F. Such a subsequence exists unless $C_0(\|x\|)$ is constant with F-probability 1. Moreover, $D(Q_n, F_n)$ converges to the optimal mean distortion for quantizing F with N levels.

In Theorem 1, convergence of the sequence of optimal quantizers $\{Q_n\}$ cannot be asserted. However, $\{Q_n\}$ can be completely partitioned into convergent subsequences whose limits are optimal quantizers for F.

In Theorem 1, F_n can be any sequence of distribution functions. We are interested in having F_n be an estimate $\hat{F}_n$ constructed from n observations of F. The principal consideration in applying the theorem is in showing uniform integrability. For sequences of estimates this becomes

$$\lim_{n \to \infty} \int C_0(\|x-y\|) d\hat{F}_n(x) = \int C_0(\|x-y\|) dF(x) \tag{2}$$

for each $y \in \mathbb{R}^k$; for almost all training sequences. Before proceeding further, we make a useful simplification of (2). *This equation requires us to find a set of sample sequences* $(X_1, X_2, \dots)$ *on which the integrals* converge for every y in $\mathbb{R}^k$. What we will be using primarily is the Strong Law of Large Numbers, which yields the conclusion:

$$\lim_{n \to \infty} \int C_0(\|x-y\|) d\hat{F}_n(x) = \int C_0(\|x-y\|) dF(x) \tag{3}$$

for almost all training sequences; for each $y \in \mathbb{R}^k$. This last equation says that for a given y, there is a set $\Lambda(y)$ of sample sequences $(X_1, X_2, \dots)$ having probability one for which the integrals converge. Obviously, (2) implies (3). However, it is shown in the Appendix that (3) implies (2). Thus (2) and (3) are equivalent. Hence if we can show (3), we may use Theorem 1 to get the results we want. In the following examples, we will

show how this can be done in a number of situations.

Empirical Distributions. Let F be the unknown k-variate distribution which we wish to (vector) quantize. If we take n independent samples $X_1, X_2, \ldots, X_n$, then the empirical distribution function is $\hat{F}_n(x) = n^{-1}(\#X_i \leq x)$, where the inequality is taken component by component. This is a simple nonparametric estimator which by the Strong Law of Large Numbers converges setwise to F for almost all sample sequences as $n \to \infty$. Take C_0 to be nonnegative, nondecreasing, nonconstant with respect to F, and lower semi-continuous. The demonstration of uniform integrability (3) is a simple application of the Strong Law of Large Numbers. We have for each y

$$\lim_{n \to \infty} \int C_0(\|x-y\|)\,d\hat{F}_n(x) = \int C_0(\|x-y\|)\,dF \quad \text{wp1},$$

provided that the last integral is finite. According to Theorem 1, then, we may quantize the $\hat{F}_n$'s, using any available method that yields an optimal solution and be assured that the resulting quantizers Q_n converge weakly to an optimal quantizer for F. This generalizes the analysis of [2,3] to other algorithms besides the extension of Lloyd's Method I. Instead of assuming the training set to consist of independent samples, we may make an ergodic (or block ergodic) assumption. This seems to be a useful assumption regarding information sources. In the development, the Strong Law of Large Numbers can be replaced by an ergodic theorem [10] and all of the conclusions reached above will remain valid.

For the rest of the paper, we will restrict our attention to distributions having a density, which we denote by f. Also, except for a portion of the last example, we will specialize to r-th power distortions, which have been widely used and studied. In this case, a natural procedure would be to use some density estimator $\hat{f}_n$ based on a training sample $X_1, X_2, \ldots$ in the quantization algorithm. Then (3) becomes

$$\lim_{n \to \infty} \int \|x-y\|^r \hat{f}_n(x)\,dx = \int \|x-y\|^r f(x)\,dx \tag{4}$$

for almost all training sequences; for each $y \in \mathbb{R}^k$. This equation says that if $\hat{f}_n$ and f are to behave in almost the same manner using r-th power quantizers, then they should have nearly the same r-th moments about arbitrary points y. Viewed differently, the quantity

$$\int \|x-y\|^r \hat{f}_n(x)\,dx$$

is an estimator of the r-th moment about y of the density f. Theorem 1 requires this estimator to be strongly consistent, i.e., to converge almost surely to the true moment of f. In addition, of course, we want the distribution associated with $\hat{f}_n$ to converge weakly to the distribution of f. This is satisfied if we have

$$\lim_{n \to \infty} \hat{f}_n(x) = f(x) \tag{5}$$

for almost all x; for almost all training sequences. The discussion thus far can be summarized by saying that an estimator $\hat{f}_n$ based on samples can be used in a quantizer design algorithm in place of f if it satisfies the strong consistency conditions (4) and (5). The examples below illustrate this point.

Normal densities. Consider that the density f to be quantized is univariate normal with unknown mean μ and variance $\sigma^2 > 0$. Let $\mathcal{N}(x)$ denote the standard normal density. Then we may write $f(x) = \mathcal{N}((x-\mu)/\sigma)/\sigma$. Let $X_1, X_2, \ldots$ be independent, identically distributed samples from f. Strongly consistent estimates of the unknown parameters μ and σ^2 are, respectively,

the sample mean and variance, $\hat{\mu}_n$ and $\hat{\sigma}_n^2$. Let $\hat{f}_n$ be a normal density with these parameters. This has the strong consistency property (5). Next we will show (4). By a change of variable

$$\int |x-y|^r \hat{f}_n(x)dx = \int |\hat{\sigma}_n x + \hat{\mu}_n - y|^r \mathcal{N}(x)dx.$$

Upon applying the c_r inequality [11, p.157] we have

$$|\hat{\sigma}_n x + \hat{\mu}_n - y|^r \leq c_r |\hat{\sigma}_n|^r |x|^r + c_r |\hat{\mu}_n - y|^r.$$

By the strong consistency of the estimators,

$$\lim_{n \to \infty} \int (c_r |\hat{\sigma}_n|^r |x|^r + c_r |\hat{\mu}_n - y|^r) \mathcal{N}(x)dx = \int (c_r \sigma^r |x|^r + c_r |\mu-y|^r) \mathcal{N}(x)dx \quad \text{wp1}.$$

Thus (4) follows upon invoking a generalized Dominated Convergence Theorem [12, p.89].

The family of normal densities in this example can be replaced with any class of almost everywhere continuous densities parametrized by location and/or scale parameters. We may contemplate other instances in which the unknown parameter is neither a location nor scale parameter, but where the above analysis can be useful. For instance, the exponent p in the generalized Gaussian density $f(x) = K_p \exp(-\gamma |x|^p)$ can be varied to fit many histograms.

In some situations it might be more appropriate to use a nonparametric density estimator. A popular type of estimator is the kernel density estimator introduced by Parzen [13] for univariate densities and generalized to multivariate densities by later authors: $f_n(x) = n^{-1} \sum_{i=1}^{n} h_n^{-k} K((x-X_i)/h_n)$. The kernel $K(x)$ is a probability density function on $\mathbb{R}^k$ and $\{h_n\}$ is a sequence of numbers decreasing to zero. Nadaraya [14] shows for the univariate case that if

> $f(x)$ is a uniformly continuous density,
>
> $K(x)$ has bounded variation, and (6)
>
> $\sum_{n=1}^{\infty} \exp(-\gamma n h_n^2)$ converges for every $\gamma > 0$,

then $\hat{f}_n(x) \to f(x)$ uniformly with probability one. The extensions of the result to several dimensions use slightly different sets of assumptions. We give the result of Moore and Yackel [15] as a typical example. If

> $f(x)$ is a uniformly continuous density on $\mathbb{R}^k$,
>
> $K(x)$ is a bounded density on $\mathbb{R}^k$,
>
> $K(x)$ has bounded variation, and (7)
>
> $n h_n^{2k}/\log n \to \infty$ as $n \to \infty$,

then $\hat{f}_n(x) \to f(x)$ uniformly with probability one.

Assuming that (6) or (7) holds, the distribution of $\hat{f}_n$ converges setwise to that of f. In addition, K and $\{h_n\}$ should be chosen appropriately so that (4) is satisfied. Two examples are given below. The first is for scalar quantization and the second for vector quantization.

Compact support. Assume that the unknown density f has compact support (which we may take without loss of generality to be contained in

[-1,1]) and is uniformly continuous on the real line. We wish to quantize this optimally using a cost function C_0 which is nonnegative, nondecreasing and lower semicontinuous, so we also assume that $\int C_0(|x-y|) K(x)\,dx < \infty$ for all y. For convenience we take the kernel to be symmetric and unimodal, so it decreases away from the origin. Consider

$$\int_{|x|>a} C_0(|x-y|)\hat{f}_n(x)\,dx \le 2 \int_{x>a} C_0(|x-y|) \frac{1}{h_n} K\left(\frac{x-1}{h_n}\right) dx \qquad \text{wp1.}$$

After a change of variable and simplification, we have

$$\int_{|x|>a} C_0(|x-y|)\hat{f}_n(x)\,dx \le 2 \int_{x>a/h_1} C_0(|h_1 x + 1 - y|) K(x)\,dx \qquad \text{wp1.}$$

It follows that $C_0(|x-y|)$ is uniformly integrable with respect to $\{\hat{f}_n(x)\}$, wp1. Therefore the kernel estimator is a viable basis for designing a scalar quantizer for a density with compact support.

Noncompact support. We can extend the previous analysis to multivariate densities with unbounded support if we restrict attention to r-th power quantizers. A bounded continuous density $f(x)$ whose tails go to 0 as $\|x\| \to \infty$ is uniformly continuous, so none of the common densities are excluded by the assumption (7). Let the cost function be $C_0(t) = t^r$ where temporarily $r = 2p$ is an even integer. We will also make the natural assumptions that $\|x\|$ has finite $(2p)$-th moments with respect to both densities f and K. Our immediate goal is to show (4) for $r = 2p$. By the c_r-inequality [11, p.157] we have

$$\|x-y\|^{2p}\hat{f}_n(x) \le c_p \sum_{j=1}^{k} n^{-1} \sum_{i=1}^{n} \left(x^{(j)}-y^{(j)}\right)^{2p} K\left(\frac{x-X_i}{h_n}\right) h_n^{-k}, \qquad (8)$$

where the superscript denotes the j-th component of the k-dimensional vector. Note that the right-hand side converges pointwise as $n \to \infty$ to

$$c_p \sum_{j=1}^{k} \left(x^{(j)}-y^{(j)}\right)^{2p} f(x).$$

If the right-hand side of (8) were integrated, we get k terms, a typical one of which looks like the following if we ignore the constant c_p:

$$m_{j,n} = n^{-1} \sum_{i=1}^{n} \int_{\mathbb{R}^k}\left[h_n x^{(j)} + X_i^{(j)} - y^{(j)}\right]^{2p} K(x)\,dx.$$

We can use the binomial theorem to expand the last integrand. After rearranging the sums we get

$$m_{j,n} = n^{-1} \sum_{i=1}^{n} \left(X_i^{(j)}-y^{(j)}\right)^{2p} + \sum_{\ell=0}^{2p-1} \binom{2p}{\ell} h_n^{2p-\ell}$$

$$\cdot \left\{ n^{-1} \sum_{i=1}^{n} \left(X_i^{(j)}-y^{(j)}\right)^{\ell} \int_{\mathbb{R}^k} [x^{(j)}]^{2p-\ell} K(x)\,dx \right\}.$$

The terms in the summation from $\ell = 0$ to $\ell = 2p-1$ are all multiplied by a power of h_n, which decreases to 0, so that in the limit, the second group of terms above is zero. Using the Strong Law of Large Numbers on the first term gives

$$\lim_{n \to \infty} m_{j,n} = \int_{\mathbb{R}^k} \left(x^{(j)}-y^{(j)}\right)^{2p} f(x)\,dx \qquad \text{wp1.}$$

This is true for all j in (8). Therefore

$$\lim_{n \to \infty} \int c_p \sum_{j=1}^{k} \left(x^{(j)} - y^{(j)}\right)^{2p} \hat{f}_n(x)dx = \int c_p \sum_{j=1}^{k} \left(x^{(j)} - y^{(j)}\right)^{2p} f(x)dx \qquad \text{wp1.}$$

Since $||x-y||^{2p} \hat{f}_n(x)$ converges to $||x-y||^{2p} f(x)$ wp1, we may use a generalization of the Dominated Convergence Theorem [12, p.89] to conclude that (4) holds for $r = 2p$. Therefore we have uniform integrability for the cost function with r an even integer. In general, for any positive r, we can let 2p be the smallest even integer for which $r \le 2p$. Then the above implies that [16, p.252]

$$\lim_{n \to \infty} \int ||x-y||^r \hat{f}_n(x)dx = \int ||x-y||^r f(x)dx \qquad \text{wp1.} \qquad (9)$$

Thus the kernel estimator $\hat{f}_n$ can be used to design a quantizer for f.

Another extension of the analysis is possible. Consider a cost function $C_0(t)$ which grows polynomially fast, i.e., there exist α, β and r so that

$$0 \le C_0(t) \le \alpha t^r + \beta. \qquad (10)$$

Since

$$\lim_{n \to \infty} C_0(||x-y||) \hat{f}_n(x) = C_0(||x-y||) f(x) \qquad \text{wp1,}$$

(9) and (10) imply, through the generalization of the Dominated Convergence Theorem [12, p.89], that

$$\lim_{n \to \infty} \int C_0(||x-y||) \hat{f}_n(x)dx = \int C_0(||x-y||) f(x)dx \qquad \text{wp1.}$$

So if C_0 is nondecreasing and lower semicontinuous, then Theorem 1 also applies, and $\hat{f}_n$ can be used to design a quantizer with the cost function C_0. Eq. (10) is equivalent to having a $\lambda > 0$ so that

$$C_0(2t) \le \lambda C_0(t) \qquad \text{for all t.} \qquad (11)$$

This is a useful form for a cost function because (11) and $\int C_0(||x||)dF < \infty$ imply that $\int C_0(||x-y||)dF < \infty$ for all vectors y.

As an application of this example, consider Max's algorithm [4] (also called Lloyd's Method II [1] or the Lloyd-Max algorithm). Some aspects of this algorithm were mentioned earlier, and we noted that as it has been described in the literature, the algorithm requires the existence of a known density function. It does not seem to be readily adapted for quantizing an empirical distribution function. However, with the approach outlined in this paper, it is possible to use density estimates as the input to Max's algorithm and get strongly consistent estimates of the optimal output levels and breakpoints. Thus the algorithm can be (indirectly) driven by a training sequence of independent observations.

APPENDIX

In this appendix we show that (3) implies (2). Eq. (3) says that for a given $y \in \mathbb{R}^k$, there is a set $\Lambda(y)$ of sample sequences $(X_1, X_2, \dots)$ having probability one for which the integrals converge. Since $\mathbb{R}^k$ is separable,

it is possible to find a dense and countable set $\{y_j\}$ for which this convergence holds with probability one. In fact, for any countable subset of $\mathbb{R}^k$,

$$\Lambda = \bigcap_{j=1}^{\infty} \Lambda(y_j)$$

is a set of probability one. Now, for a given y in (2), we may find a finite subset A of the y_j's so that

$$C_0(\|x-y\|) \leq \sum_{y_j \in A} C_0(\|x-y_j\|).$$

This involves constructing a cube around y and picking the y_j outside the cube so that $\|x-y\| \leq \|x-y_j\|$ for at least one y_j. Then

$$\sup_n \int_{\|x\|>a} C_0(\|x-y\|)\,d\hat{F}_n(x)$$

$$\leq \sum_{y_j \in A} \sup_n \int_{\|x\|>a} C_0(\|x-y_j\|)\,d\hat{F}_n(x).$$

For each y_j and each sample sequence in Λ, the right-hand side can be made arbitrarily small by appropriately choosing a. Therefore $C_0(\|x-y\|)$ is uniformly integrable for every y, with respect to every $\{\hat{F}_n\}$ arising from the set Λ. This is equivalent to (2). Therefore, (3) implies (2).

ACKNOWLEDGEMENT

REFERENCES

[1] S. P. Lloyd, "Least squares quantization in PCM," Bell Telephone Labs Memorandum, Murray Hill, NJ, 1957, reprinted in *IEEE Trans. Information Theory*, IT-28, no. 2, pp. 129-137, March 1982.

[2] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communication*, COM-28, no. 1, pp. 84-95, Jan. 1980.

[3] R. M. Gray, J. C. Kieffer and Y. Linde, "Locally optimal block quantizer design," *Information Control*, 45, no. 2, pp. 178-198, May 1980.

[4] J. Max, "Quantizing for minimum distortion," *IRE Trans. Information Theory*, IT-6, no. 1, pp. 7-12, March 1960.

[5] D. K. Sharma, "Design of absolutely optimal quantizers for a wide class of distortion measures," *IEEE Trans. Information Theory*, IT-24, no. 6, pp. 693-702, Nov. 1978.

[6] E. F. Abaya and G. L. Wise, "Some notes on optimal quantization," *Proc. International Conference on Communications*, vol. 2, pp. 30.7.1-30.7.5, June 14-17, 1981.

[7] E. F. Abaya and G. L. Wise, "On the existence of optimal quantizers," *IEEE Trans. Information Theory*, IT-28, pp. 937-940, Nov. 1982.

[8] E. F. Abaya and G. L. Wise, "Convergence of vector quantizers with application to the design of optimal quantizers," *Proc. 19th Annual Allerton Conference on Communication, Control, and Computing*, pp. 79-88, Sept. 30-Oct. 2, 1981.

[9] E. F. Abaya and G. L. Wise, "Convergence of vector quantizers with applications to optimal quantization," *SIAM J. Applied Mathematics*, 44, pp. 183-189, Feb. 1984.

[10] L. Breiman, "The individual ergodic theorem of information theory," *Ann. Math. Stat.*, 28, pp. 809-811, Sept. 1957.

[11] M. Loeve, *Probability Theory 1*, 4th Edition, Springer-Verlag, New York, 1977.

[12] H. L. Royden, *Real Analysis*, 2nd Ed., Macmillan Publ. Co., New York, 1968.

[13] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, 33, pp. 1065-1076, Sept. 1962.

[14] E. A. Nadaraya, "On non-parametric estimates of density functions and regression curves," *Th. Prob. Appl.*, vol. X, pp. 186-190, 1965.

[15] D. S. Moore and J. W. Yackel, "Consistency properties of nearest neighbor density function estimators," *Ann. Stat.*, 5, pp. 143-154, Jan. 1977.

[16] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, 2nd Ed., John Wiley and Sons, New York, 1971.

END

FILMED

11-85

DTIC